

DOCUMENT RESUME

ED 184 818

SE 030 349

AUTHOR Blosser, Patricia E.
 TITLE Review of Research: Teacher Questioning Behavior in Science Classrooms.
 INSTITUTION ERIC Information Analysis Center for Science, Mathematics, and Environmental Education, Columbus, Ohio.
 SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
 PUB DATE Dec 79
 CONTRACT 400-78-0004
 NOTE 52p.
 AVAILABLE FROM Information Preference Center (ERIC/IRC), The Ohio State University, 1200 Chambers Rd., 3rd Floor, Columbus, OH 43212 (\$1.75).
 EDRS PRICE MF01/PC03 Plus Postage.
 DESCRIPTORS Educational Research: Elementary Secondary Education: Higher Education: Inservice Education: Preservice Teacher Education: *Questioning Techniques: *Research Reports: *Science Education: *Science Teachers: *Teacher Behavior

ABSTRACT

Selected for this review are dissertations and other research reports related to science teacher questioning behavior, with particular emphasis on those studies designed to help teachers change their questioning behavior. Summarizing the section on observational studies (N=11) the author concludes that science teachers appear to function primarily at the level of cognitive-memory thinking operations in their questioning behavior, regardless of educational level taught. The section on experimental studies clusters research for preservice (1) elementary teachers (N=11), (2) secondary teachers (N=9), and (3) college teachers (N=2). It appears more possible to help preservice elementary teachers than secondary or college teachers improve their questioning behavior. Reviews for inservice elementary (N=3) and secondary (N=3) teachers indicate that elementary school science teachers' questioning behavior can change, in some instances, and that different points of view relative to questioning are provided by the research dealing with secondary school science teachers. Also reviewed are four experimental studies focusing on wait-time. An additional section reviews the research, clustered by question category systems used, length of instructional treatment, materials used, and outcomes of the studies. Needs for further research are projected. (CS)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED184818

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

REVIEW OF RESEARCH

TEACHER QUESTIONING BEHAVIOR IN SCIENCE CLASSROOMS

Patricia E. Blosser
Faculty of Science & Mathematics Education
The Ohio State University



Clearinghouse for Science, Mathematics
and Environmental Education
The Ohio State University
College of Education
1200 Chambers Road, Third Floor
Columbus, Ohio 43212

December 1979

Research time for this paper was supported in part by the University Small Research Grant Program, Graduate School, The Ohio State University, Columbus, Ohio.

SE 030 349

SCIENCE EDUCATION INFORMATION REPORTS

Science Education Information Reports are issued to analyze and summarize information related to the teaching and learning of science education. It is hoped that these publications will provide information for personnel involved in development, ideas for teachers, and indications of trends in science education.

Your comments and suggestions for this series are invited.

Stanley L. Helgeson
Associate Director, Science Education

Patricia E. Blosser
Faculty Research Associate, Science
Education

Sponsored by the Education Resources Information Center of The Ohio State University.



This publication was prepared with funding from the National Institute of Education, U.S. Department of Health, Education, and Welfare under contract no. 400-78-0004. The opinions expressed in this report do not necessarily reflect the positions or policies of NIE or PEW.

TABLE OF CONTENTS

INTRODUCTION	1
Sources of Studies Reviewed	2
OBSERVATIONAL STUDIES	3
EXPERIMENTAL STUDIES	7
Experimental Studies Involving Preservice Elementary School Teachers	7
Experimental Studies Involving Preservice Secondary School Science Teachers	14
Experimental Studies Involving Preservice College Science Teachers	23
Experimental Studies Involving Inservice Elementary School Science Teachers	25
Experimental Studies Involving Inservice Secondary School Science Teachers	28
Experimental Studies Focused on Wait-Time	30
ADDITIONAL CONSIDERATIONS	35
Question Category Systems Used	35
Length of Instructional Treatment	37
Materials Used in Treatment	38
Outcomes of Studies	39
IN CONCLUSION	42
LITERATURE CITED	43

REVIEW OF RESEARCH:
TEACHER QUESTIONING BEHAVIOR IN SCIENCE CLASSROOMS

INTRODUCTION

The statement that "little has been done" is frequently used in literature reviews relative to a particular research topic, sometimes with more justification than lack of it. Such a statement ". . . However, very little has been done in preservice teacher training to provide the preservice (science) teacher with instruction on question-asking skills and strategies. . . ." (Rice, 1977, p.358) raised the issue of how many studies must be identified before the "very little" designation no longer holds true.

Since the late 1960's the author has been assembling a file of dissertation abstracts and other research reports related to teacher questioning. The amount of material in this file is evidence that much has been written about the ways teachers ask questions.

A large number of these studies involve teachers in areas other than science and include research with elementary classes in which lesson content is not specified. Some of the research is descriptive; some, experimental. There are also studies of children's questions, questions included in textual material, and examination questions. Again, some of these studies involve science content but some do not.

For the purposes of this review, only those research studies involving science teachers (preservice, inservice) were analyzed. Particular emphasis was placed on those studies designed to help teachers change their questioning behavior.

The questions teachers ask have long been considered important. This holds true regardless of the educational level being considered or the content. Hilda Taba's work on thinking in elementary school children contains this quote:

The examination of the tapescripts. . . leaves no doubt that the nature of the questions has a singular impact on the progression of thought in the class. The questions teachers ask set the limits within which students can operate and the expectations regarding the level of cognitive operations. Questions are the carriers of whatever new cognitive system is emerging. (Taba in Stewart, 1975, p. 15).

In a chronological review of observational studies of teacher questioning, spanning 1893-1963, Hoetker and Ahlbrand (1968) report that one observer commented that it appeared the central focus of all of the questioning activity was the textbook and that the teacher's job appeared to be to see that the students studied it (p.5). With the development of science content improvement projects and the emphasis on inquiry as a teaching technique, some researchers have attempted to change science teachers' questioning behavior to promote more student inquiry and less

teacher direction in science classes. The rationale for wishing to change teachers' questioning patterns appears to be that, while factual information is needed as a basis for structuring investigations, teachers need to ask a variety of questions designed to promote more than factual-recall thinking activities by their pupils.

In addition to attempting to change the types of questions teachers ask, a related line of research has involved helping teachers learn to wait after asking a question or accepting a student response to give pupils time to think and to expand upon their responses. The usual wait-time or pause is a short one and the behavior change desired is that of lengthening the pause.

Sources of Studies Reviewed

Studies were identified from the file mentioned earlier in this paper, from scanning recent issues of Dissertation Abstracts International from a search of the ERIC data base (Resources in Education, Current Index to Journals in Education), and from the Reader's Guide. The time period emphasized was that of the late 1950's through the present (although studies earlier than 1955 were not excluded), to parallel the use of science course improvement projects and the emphasis on inquiry in science education.

Whenever possible, the document rather than its abstract or resume was read. The majority of the research identified consisted of doctoral dissertations. Seven of these dissertations had been completed by persons enrolled at Indiana University. This common source would seem to indicate either an ongoing line of faculty research or strong support by one or more science education faculty members. Because of the variety of approaches involved in these seven dissertations and the apparent lack of building upon previously conducted research (although the studies were usually cited in the literature review chapter), the second guess would appear to be the more correct one.

In addition to the separate studies, there are available some documents in which a number of studies are briefly identified and discussed. In 1973, Beisenherz and Tucker presented a paper at the annual meeting of the National Association for Research in Science Teaching in which they talked about limitations of research in science teacher questioning behavior. Also published in 1973 was A Review of Research on Teacher Behavior (Balzer, et al.) in which science teacher questioning behavior research was described. In 1978 in Volume 1 of What Research Says to the Science Teacher, edited by M. B. Rowe, Chapter Two contains a discussion of questioning behaviors of science teachers. Although several sources of information are available, the two 1973 documents do not include recent research and the 1978 publication does not focus solely on science teachers' questions although the chapter title implies this but includes research from other subject areas as well. This review will contain a brief discussion of observational studies of science teachers' patterns of question asking and will concentrate on those investigations emphasizing attempts to change questioning behaviors.

OBSERVATIONAL STUDIES

Before setting out to change a situation, it is wise to determine what conditions presently exist. Observational research can help provide data. Early research on questioning was of the observational variety rather than experimental. Research on questioning in science classrooms also began with observational techniques.

Kleinman's dissertation research (1964) took the form of an exploratory study of questioning practices of general science teachers. She was interested in identifying the kinds of questions asked and investigating the relationship, if any, between the kinds of questions asked and teacher and pupil behaviors. A third problem in Kleinman's study was to determine whether the kinds of questions general science teachers asked influenced pupils' understanding of science. She found (1) the kinds of questions teachers ask are fairly stable for each teacher, (2) teachers who ask more critical thinking questions also tend to ask fewer questions per minute, (3) teachers who ask more critical thinking questions also ask more neutral, clarifying, and associative questions and fewer rhetorical or factual questions to a degree significant at the 0.01 level of confidence for each category of questions, and (4) only one value question was asked in all 35 observations (1964:99-107).

A year later Moyer published the results of an exploratory study of the instructional processes in selected elementary schools, using data from 14 science lessons which had been observed and tape-recorded (1965). Moyer concentrated on the questions identified in these science lessons, analyzing the questions according to (1) type, (2) structural form, (3) function, (4) relationship between structure and function, (5) teacher development and utilization of questions, and (6) teacher awareness of the questioning process.

Moyer compiled 2,500 questions from the 14 science lessons. He found that none of these questions stimulated an evaluative response from the pupils involved. Two-fifths of the questions required the pupils to respond in ways requiring little or no mental effort. The measure of questioning effectiveness, Moyer decided, seemed to depend on the importance of the function, percentage of responses, and quality of content. Noting that many inadequate responses were accepted by teachers, Moyer concluded that teachers are not prepared to develop and effectively utilize the questioning process.

Some observational studies could be clustered because, although carried out at different institutions, all involved inservice elementary school teachers and the use of the Science Curriculum Improvement Study (SCIS) materials (Bruce, 1969; Garigliano, 1973; Kondo, 1969; Moon, 1969; Porterfield, 1969; Rowe, 1974a; Wilson, 1969).

Bruce (1969) examined the extent of the relationship among selected teacher personality factors, science process skills, attitude toward teacher-pupil relationship, and the verbal characteristics of question asking. Thirty-three teachers participated, with 15 being observed and

tape recorded prior to participation in a three-week SCIS workshop. Questions from 220 science lessons were analyzed, using Bloom's Taxonomy. Bruce found a significant difference in the level of questions asked before and during formal involvement in the SCIS program. The number of high level questions increased, with a significantly greater proportion being analysis questions.

Garigliano's study (1973) was one of four investigations relating to wait-time and will be discussed later in this review.

Kondo (1969) studied the classroom behavior of SCIS teachers to determine if a possible relationship existed between questioning behavior and different types of SCIS lessons (invention, discovery). Using four first-grade teachers in the same school, Kondo found that the way the lesson was approached (teacher demonstration vs. children handling materials) had a greater influence on the types of questions asked by the teachers than did the type of lesson (invention, discovery).

Moon (1969) attempted to analyze selected examples of verbal behavior patterns in primary grade classrooms (N=32) during science activities. Sixteen teachers used SCIS materials. After the SCIS teachers had participated in a workshop designed to help them use SCIS materials, the question preference changed from low order questions to high level questions.

Porterfield's study (1969) involved a slight variation in that he was interested in seeing if preparation in the use of SCIS materials influenced questions asked during reading instruction. Porterfield made audiotapes of reading lessons involving 32 teachers (16 for grade 2 and 16 for grade 4). Half of the teachers at each grade level used SCIS and half did not.

Porterfield analyzed the questions from the reading lessons (two per teacher, taped one week apart) using Harris and McIntyre's Teacher Question Inventory (one of the many variations on Bloom's taxonomy--includes two question categories of an affective nature). He concluded that questioning skills from SCIS science lessons did appear to transfer to reading instruction in that SCIS-trained teachers showed significant differences in using questions related to the higher cognitive processes (translation, interpretation, analysis, synthesis). However, recall-type questions were used more frequently than any other question type by all of the teacher groups.

Rowe's data (1974a), although resulting from classrooms in which SCIS materials were used, focused on wait-times and reward schedules and also will be discussed later in this review.

Wilson (1969) analyzed the teaching procedures of 30 elementary teachers, 15 of whom were using SCIS materials. When Wilson analyzed the questions asked in science lessons which had been tape-recorded, he found that recognition and recall questions were asked a significantly greater proportion of the time by teachers using the more traditional elementary school science materials. SCIS teachers asked

a significantly larger proportion of analysis and synthesis questions as well as asking more questions than did teachers using traditional materials for science.

Hunter (1969) analyzed the verbal behavior of 22 first-grade teachers as they taught science, using the Revised VICS-Science observation system. She found no significant difference in the amount of questioning behavior between the group of 11 teachers who had participated in an inservice program related to the elementary science curriculum improvement materials and those with no special inservice training. Hunter stated that, although some of the teachers were using the newer science materials, teacher questions were not stimulating divergent and evaluative thinking activities. Changes in curriculum content did not necessarily result in changing teacher behavior.

Larson (1973) conducted a study designed to describe the questioning strategies of fifth-grade teachers when they taught science lessons. Larson investigated the initial questioning interchange made by the teacher and classified the questions using the Questioning Strategies Observation System (QSOS). He also considered the form of student response elicited and the teacher's reaction to the student response. Nineteen teachers, with 1 to 46 years of teaching experience, were observed.

Larson's category system contains four major headings: teacher-eliciting statements, cognitive content categorizations (knowledge, convergent thinking, divergent thinking, evaluation, affectivity, procedure), student responses, and teacher reaction to response.

Larson's report contains data on each of the 19 teachers involved. These data were obtained from the analysis of 20 minutes of interaction from audiotapes of observations ranging from 16 to 42 minutes in length. Three judges analyzed the tapes along with Larson. Rater reliability was determined by use of a correlation coefficient for each of the three tapes analyzed (.83, .92, .96).

Larson found that teachers tended to ask questions at the simple recall level. Teachers also tended to ask questions without any designation of who should respond and they tended to be positive in their reactions to student responses.

Based on these observational studies, it appears that science teachers function primarily at the level of cognitive-memory thinking operations in their questioning behavior regardless of educational level involved (elementary or secondary). None of the studies reviewed was concerned with making a direct improvement on teacher questioning as a part of the research but only with observing what took place during science lessons.

EXPERIMENTAL STUDIES

The studies discussed in this section will be grouped on the basis of whether the teachers involved were preservice or inservice teachers. Not only do these two groups differ relative to amount of teaching experience, they may also differ in manner of involvement in the research. Preservice teachers frequently become participants in a study because they happened to enroll in a science methods course during a particular quarter or semester. They may be randomly assigned to a treatment or to a control group but they seldom volunteer to participate in the research. The use of volunteers is more frequently reported when inservice teachers are involved.

Do variables of age, experience, and willingness to participate make a difference? Some persons think they do. Beisenherz and Tucker, however, say "... the effect of the manipulation of teacher age, experience, and content background variables on the improvement of classroom questioning is vague at best..." (1973, p. 2). Nevertheless, preservice and inservice teachers are grouped separately for discussion purposes in this research review.

Table 1

Experimental Studies of Questioning Behavior Involving
Preservice Science Teachers

ELEMENTARY

Anshutz, 1973
Bozardt, 1974
Cunningham, 1968
Doyle, 1972
Grunau, 1972
Koran, 1969, 1970
Masla, 1968
Rice, 1977
Riley, 1977
Santiesteban, 1974

SECONDARY

Adhikary, 1973
Blosser, 1970, 1972
Boeck, 1970
Coyne, 1974
Crawley & Krockover, 1979
Konetski, 1969
Lamb, 1977
Smith, W. S., 1973

COLLEGE

Murphy, 1972
Smith, A. B., 1977

Experimental Studies Involving Preservice
Elementary School Teachers

Although Cunningham considered that his study was descriptive rather than experimental, it has been placed in the "experimental studies" group because Cunningham did attempt to modify questioning behavior. (Cunningham says, page 7 in his dissertation, that he considers his study to be descriptive because no comparison or control groups were used (1968).) Cunningham considered that phrasing questions was a major factor in identifying the relevancy of questions. He sought evidence that a method of instruction could bring about changes in question-phrasing ability. Forty preservice teachers (p. 121, 1968, reports an N of 38) were involved in a seven-day instructional sequence. When data from pre- and posttests were compared, Cunningham found that instruction appeared to have significantly reduced the number of cognitive-memory questions asked and increased the number of divergent thinking questions recorded. No significant difference in numbers of convergent thinking questions was identified. No evaluative questions were identified in either pre- or posttest tapes.

Masla's study focused on the use of interaction analysis instruction and its effect on the verbal inquiry patterns of preservice elementary teachers. One category of the interaction analysis system dealt with questions and was subdivided into cognitive memory questions, convergent thinking questions, divergent thinking questions, and evaluative thinking questions. (The basis for this subdivision as well as for the categories in Cunningham's study is the system developed by Gallagher and Aschner and which was based on the major divisions of the Guilford structure of the intellect model.)

The preservice students (N=76) were pretested with the Elementary Teacher's Science Inventory (ETSI) to determine competency in science processes, assigned to a high competency or a low competency group on the basis of their ranked scores, and randomly assigned by rank to an experimental group or a control group. The experimental group received intensive instruction in interaction analysis. After instruction, 40 students from the total group were randomly selected to teach science lessons to elementary school children who had volunteered to participate in the study. A significant effect attributable to differences between the means of the question ratios of the two groups was found, with the direction of significance in favor of the experimental group. Experimental group means indicated a significantly greater proportion of open-ended questions. Level of competency in science processes did not appear to be a factor affecting the verbal inquiry patterns of the teachers involved.

The two reports by Koran relate to the use of models in teacher education. In the 1969 paper Koran reported about an investigation to test the effects of telling teachers how to teach as compared to the effects of presenting teachers with visual or film-mediated models of the behavior to be acquired on teachers' performance under simulated conditions. The behavior emphasized was asking observation and classification questions. Students enrolled in an elementary science methods

course were randomly assigned to two treatment groups or a control group. All three groups were asked to generate as many observation and classification questions as possible, using kit materials from Science--A Process Approach (trial 1). Then treatment group I viewed a 14-minute videotaped model of a teacher conducting a science lesson with four elementary school children. In this model, the teacher's observation and classification questions were highlighted. Treatment group II and the control group received a placebo during this time. Trial 2 consisted of again generating written questions, as did trial 3 which took place after treatment group I again viewed the filmed model. Koran found that students who had viewed the videotaped model of questioning scored significantly higher on both within-group and between-group differences than did the two control groups (specific instruction but no filmed model; no treatment) (1969, p. 222).

The 1970 paper by Koran continued this line of research. Film-mediated models appear valuable for teacher education purposes but are expensive. Koran was interested in identifying an equally or more effective method and also possibly one that was less expensive and more portable. He decided to compare the relative effects of self-rating and a film-mediated model on the production of observation and classification questions. In the self-rating situation the trainees attempted to generate the criterion behavior prior to training and then rated their performance as a part of the training, using a rater protocol which identified the correct responses. Students in the film-mediated situation observed a filmed performance of a teacher displaying the behaviors to be acquired.

As in the 1969 research, all students had three trials at writing observation and classification questions. A retention test was given two weeks after the third trial. Koran concluded that both training methods were equally effective with those subjects and on the criterion task. No significant differences were found for retention test results. However, Koran reported that he could find evidence in the data to suggest that the self-rating group tended to lose proficiency after two weeks "... at almost a significant level ($p < .10$)" (1970, p. 387).

Doyle (1972) emphasized training in the use of probing questions--those questions which cause the person responding to go beyond the initial response. Probing questions, in Doyle's study, were defined as asking for clarification, verification or justification, or the extension of knowledge. Doyle was interested in learning if preservice teachers could learn to use probing questions and, if so, if there was a minimal amount of instruction that would produce maximal results. Fifty-four preservice elementary teachers were involved in a three-phase instructional program and were placed in three groups of 18 students each, with each group receiving different amounts of instruction. (Initially it was reported that 55 teachers were involved, 1 male and 54 females, but one subject was randomly eliminated, to produce equal N's.)

Doyle found that instruction did help the elementary education students write and classify probing questions, write appropriate probing

questions in response to written pupil replies as well as in response to pupils' verbal (oral?) replies. However, there was no significant gain in the use of probing questions when pre- and posttest measures were compared. Doyle speculated that his subjects were already approaching the level of optimum use of probing questions before they received instruction. The finding that only the group with the lowest mean on the pretest showed any appreciable gain on the posttest provided further support for this speculation.

Grunau's study (1972) is included in this review although its primary focus was not on questioning skill development but upon preparing preservice elementary teachers to teach selected processes of science to children. Grunau used two instructional methods in his study. Method one was designed to improve the skills of observing, classifying and inferring in the preservice teachers while method two focused on eliciting observing, classifying, and inferring behavior in children by the use of questions. Hall's Instrument for the Analysis of Science Teaching (IAST) and Fischler and Anastasiow's Science Teaching Observation Instrument (STOI) were used to analyze the verbal interaction. Grunau reported finding no significant difference in the effects of the two instructional treatments. However, teachers of high initial process skill (as measured by Sweetser's Elementary School Teacher's Science Process Test), undergoing either treatment, demonstrated teaching behaviors significantly different from those of low initial process skill. The difference was due in large part to question complexity level, with higher initial process skill levels associated with a higher final proportion of observing, classifying, and inferring questions asked.

Anshutz (1973) considered that inquiry teaching was little used or poorly understood and that, as a part of learning to use inquiry, teachers needed to modify their questioning behavior and the use of wait-time. She investigated the effects of training in questioning skills and wait-time on the subsequent classroom behavior of preservice elementary teachers. Twenty-six students were randomly selected from two intact methods classes and assigned to an experimental or a control group (13 in each).

The methods class involved both on-campus and field work. Micro-teaching with peers was done as a part of the on-campus activities. Prior to the field experience there were no significant differences between rankings of mean wait-time and percentage of a particular kind of question when data from the microteaching lessons of experimental and control groups were compared. During the four weeks of field experience, nothing was done to the experimental group students to reinforce wait-time behavior or questioning techniques.

Anshutz reported that, at the end of the four weeks of field work, the experimental group asked more questions than did the control group and had a composite wait-time twice that of the control. Questions identified on audiotapes were categorized using Parson's Self-Guided Analysis System which classified questions as rhetorical, informational, leading, or probing. Anshutz did the bulk of the audiotape analysis and

did not determine rater reliability "... because tallies agreed so much..." (1973, p. 48). Even when a second observer was involved in the study, no rater reliability determination was attempted. Anshutz looked only at the number of tallies per question category, not at a question-by-question comparison of categories.

Santiesteban (1974) completed an investigation that is related to the studies of Koran reported earlier in this paper. Santiesteban compared the relative effects of videotape and audio models on the acquisition of observation and classification skills by preservice elementary teachers. One group of students watched a videotape model of a teacher using observation and classification behaviors. Another group (audio model) was given only the audio track of the videotape. A control group received neither model.

Treatment, as reported in Santiesteban's study, was relatively brief. Students received five minutes of set induction in which the concept, the procedures, and the materials were introduced. The video and audio models each ran ten minutes. Students were given five minutes to prepare a lesson to be taught in a microteaching situation of fifteen minutes in length. Following the microteaching lesson, teachers and their pupils took some tests--with this testing requiring approximately 20 minutes. Total time for the preservice teachers was 55 minutes.

Three raters, each trained for 20 hours, performed the analysis. The method for determining rater reliability was not given but level of reliability for observation questions was reported as .92 and for classification questions, as .92 also. A significant difference ($p < .05$) was found between the audio model group and the control group on the teacher criterion test, in favor of the audio model. There were no significant differences between the video model and the control groups or between the video and audio model groups.

Bozardt (1974) reported that a review of the literature indicated that teachers need to use more higher level, thought-provoking questions in the classroom to help their students develop their abilities to think. She designed a study to determine whether instruction in asking questions systematically resulted in a greater use of questions other than those at the memory level. Thirty-eight students from three sections of an elementary science methods course were involved in the study.

Bozardt used Sanders' question hierarchy in the questioning modules she developed. These modules included reading materials, activities, and self-evaluation activities. Subjects worked through the modules individually and at their own pace.

The three methods sections were used in a randomized block design for Bozardt's study. Each class section was assigned to a different portal school for field work and each section contained experimental and control group members. All methods students taught a sequence of six lessons in the school. All received a preliminary module. The students in the experimental group received the module on questioning after their first taping. The fifth lesson taught by all students was

chosen for data analysis by the investigator and two judges. Rater reliability as determined by Ebel's intraclass correlation coefficient was .69 for the experimental group, .54 for the control group, and .61 for all tapes.

Bozardt reported no significant differences in the proportion of questions at levels higher than memory for preservice teachers receiving instruction in questioning and those not receiving instruction. Neither were there significant differences in the proportion of incidents which were systematic between the two groups nor in the amount of time spent at levels higher than memory in the questions asked by the two groups. Bozardt reported that the largest single proportion of questions asked was in the noncognitive category and that, in the cognitive category, the largest proportion of questions asked were those requiring memory for a response.

Rice (1977) hypothesized that many newer instructional programs in elementary science depend on the ability of teachers to ask thought-provoking, inquiry-type questions and that improved teacher-directed questioning would be a valuable and necessary instructional tool. He decided to investigate if wait-time, the number of questions asked, and the cognitive level of questions would improve if preservice elementary teachers were given instruction dealing with various question-asking strategies. Rice randomly assigned ten students to an instructional and a noninstructional group (five in each). These ten students were currently enrolled in an elementary science methods course prior to student teaching.

The students each taught a lesson to a classroom of elementary pupils and the lesson was audiotaped for pre-treatment data concerning wait-time, number of questions asked, and cognitive level of the questions. A panel of three judges was used with interrater reliability (determined by a one-factor within-repeated-measures ANOVA) of 0.89 for wait-time, 1.00 for number of questions asked, and 0.97 for cognitive level of questions. All ten preservice teachers prepared six science lessons using SCIS materials.

The instructional group viewed three films from the Far West Lab's minicourse I, Effective Questioning Elementary Level, read and discussed Rowe's (1969) "Science, Silence and Sanctions," as well as discussing the number and cognitive level of the questions each teacher asked during lessons one through five. The preservice teachers, along with the investigator, categorized the questions asked as low cognitive or high cognitive. They also read some additional material on question asking at their own rate during the following week. (The instructional treatment took place over a period of six hours during a two-week period.) Although the noninstructional group also met with the investigator to discuss the lessons they presented, they did not see the films or receive information about questioning.

Both groups made a sixth audiotape in which time of day, environmental conditions, and teaching materials were held constant with lesson one. The three judges again analyzed this audiotaped lesson with interrater reliability of 0.92 for wait-time, 0.98 for number of questions asked, and 0.88 for cognitive level of questions.

Rice reported that the instructional group showed a large increase in wait-time from pre- to post-audiotape with no such increase for the noninstructional group. The instructional group decreased the number of questions asked (by 56 percent) while the noninstructional group's number of questions increased (23 percent). The instructional group also exhibited an increase in cognitive level of questions asked while there was no such increase for the noninstructional group. Rice concluded that his data supported the hypothesis he set out to test: preservice elementary teachers receiving instruction on specific question-asking strategies could show significant improvement over preservice elementary teachers not receiving instruction (1977, p. 357).

Rice suggested that modular programs developed on question asking be included in elementary science methods courses, that microteaching lessons be video or audio-taped and teacher strategies discussed on an individual or whole class basis, and that preservice teachers should observe model teachers and have direct contact with elementary pupils in their presentation of elementary science lessons.

The most recently published study dealing with an experimental treatment focused on questioning behavior of elementary school teachers was reported by Riley (1978). Riley investigated whether learning a question classification hierarchy would result in teachers asking higher cognitive level questions in the classroom. From a group of 100 senior level preservice elementary teachers 40 were randomly selected and assigned to two treatments and to schools using different science programs. The 20 preservice teachers in the treatment group received three and one-half hours of instruction on question classification. The Question Classification System for Science was used. The students read definitions of the question categories, saw examples, and classified questions. The 20 students in the control group received no instruction.

Thirty-minute audiotapes of science lessons were recorded for analysis by three raters (rater reliability not specified). Riley found that teachers who had received the training asked significantly higher level questions than those preservice teachers with no training. The science curriculum used also had a significant effect on the cognitive level of the teacher's questions with those teachers using the Science--A Process Approach materials having a mean cognitive-level score significantly higher than those teachers using the textbook program Concepts in Science. In addition the SAPA teachers asked significantly fewer questions than did those teachers using the textbook program. There was no significant difference between treatment and control groups on the number of questions asked.

Riley's findings, like those of Rice, support the idea that training in the meaning and use of a question classification system can raise

the cognitive level of questions asked by preservice teachers. Riley's results also indicate that the curriculum used may influence the cognitive level and number of questions teachers ask.

In Summary

The eleven studies of attempts to improve the questioning behavior of preservice elementary school teachers provide, for the most part, some evidence to encourage science educators who wish to help improve questioning behavior. The use of models is helpful (Koran) whether audio or video, and apparently an audio model is sufficient for some improvement in questioning techniques (Santiesteban). If teachers have some skill in the use of science processes, this skill appears related to the complexity of questions they ask (Grunau). Experiences with elementary school children appear to be preferable to experiences with peers in developing questioning skills (Anshutz). Instruction can result in use of higher cognitive level questions as well as in increased wait-time (Rice). Preservice teachers can learn to use a category system for classifying questions (Riley). Not only does instruction in questioning influence teachers' questioning behavior so does the science curriculum which the teachers are using (Riley).

In the studies in which no significant differences were found (Doyle, Bozardt), some factors in addition to instructional treatment may need to be considered. Doyle speculated that subjects in his experimental group may have been approaching the level of optimum use of probing questions prior to instruction, his detailed instructions prior to pre-testing may have cued his subjects toward the use of probing questions in their lessons without his intending to do so, and he was unable to isolate treatment groups from each other during instruction. Doyle also mentions negative attitudes on the part of the preservice teachers toward the treatments.

Subjects in Bozardt's study worked through the questioning module individually and at their own pace. Those in the other studies appear to have had less freedom to pursue (or not pursue) instruction. Perhaps the amount of control an individual exercises relative to his/her participation plays a part in the results. One of the investigators who worked with secondary school teachers (Lamb) considers this important. His views will be discussed in the next section of this review.

Experimental Studies Involving Preservice Secondary School Science Teachers

Konetski (1969) studied the effects of two instructional strategies on the use of divergent and evaluative questions asked by preservice secondary school science teachers. He used students enrolled in three intact methods classes. All students made a pretest audiotape of a science lesson. The tapes were analyzed for divergent and evaluative questions asked.

Subjects were grouped as high or low. Equal numbers of high and low students were randomly assigned to two groups within an intact class. Two instructors and two strategies were randomly assigned to the groups. Each strategy involved three days of instruction. On day 1, students were made aware of the value of questioning and were given a programmed instruction booklet--for both strategies. The strategies diverged on day 2. For strategy 1, day 2 of instruction involved categorization of questions from an audiotape of a science lesson and a discussion of the questions asked. For strategy 2, day 2 involved designing questions for inquiry-discovery type science lessons from problem situations given to the preservice teachers. For strategy 1, day 3 involved categorization of questions from audiotaped science lessons and a discussion of questions that should be asked in science. For strategy 2, day 3 involved the evaluation and categorization of questions developed by peers and a discussion on the use of divergent and evaluative questions.

After the strategies were administered, the subjects taught a second science lesson (posttest one), had a conference with the instructor, and then taught a third lesson (posttest two). Students in the control group received "informal instruction" on questioning in the form of a handout as well as a conference with the instructor between the first and second lessons taught.

When Konetski analyzed his data he found that there was a significant effect ($p < .01$) on the number of divergent and evaluative questions after formal instruction, a significant effect ($p < .001$) on the proportion of divergent and evaluative questions after formal instruction, and a decrease, with a significant effect ($p < .01$), in the total number of questions asked after formal instruction. When the two formal strategies were compared with the informal instruction, subjects in both strategy 1 and strategy 2 asked significantly more divergent and evaluative questions, in both posttest one and posttest two, than did those in the control group. Konetski concluded that either formal strategy would work but that strategy two might be the more desirable even though there were no significant differences in the data.

Boeck (1970) designed a study to explore the usefulness of precision teaching strategies in microteaching as a means of changing behavior. "Precision teaching" derived from operant conditioning and, in Boeck's study, focused on changing the rate of high level questions asked by preservice secondary teachers. High level questions were identified as those dealing with Bloom's taxonomical levels of application, analysis, synthesis, and evaluation.

Boeck speculated that high level questions should require more thoughtful and detailed responses from students so the amount of pupil talk should increase. This speculation related to the second problem in her study: that of the relationship between teaching behaviors and desirable pupil behaviors (increased talk).

Twenty juniors enrolled in educational psychology and a science methods course were assigned to two treatment groups (seven each) and a control group (six students). Ten of the students were physical science majors and ten, biological science. These students were assigned to teach ten 10-minute lessons to junior high school students. After five lessons, seven students were randomly selected and taught to categorize and graph their own questioning behavior. After seven lessons, seven more students received this same treatment. The remaining six students served as the control group.

Boeck and another individual analyzed the data and rated the questions asked. No method of determining rater reliability was reported although Boeck said that the raters reached consensus on the categorization of each question as being either high level or low level (memory, translation). Pupil talk and teacher talk were timed by using two timers, one ran when the teacher talked and the other, when pupils talked. No timer was used for periods of silence.

Boeck reported there was no significant difference among the three groups prior to treatment in rate of high level questions. On days 8, 9 and 10 there was a significant difference, at the .05 level, between the two experimental groups and the control group in the rate of high level questions asked, in favor of the experimental groups.

There was no significant difference among the three groups prior to treatment in the rate of low level questions. There were no significant differences among the three groups on low level questions on days 8 and 9, but the experimental groups asked low level questions at a significantly ($p < .05$) higher rate than did the control group on day 10.

When teacher talk and pupil talk were considered, none of the orthogonal planned contrasts showed significance at the .05 level. However, the ratio of pupil talk increased and the ratio of teacher talk decreased for the experimental groups on days 8, 9 and 10.

It appeared that students taught to categorize and graph their own high level questioning behavior increased the rate at which they asked high level questions. There was no clear effect on low level questioning behavior. Students were not asked to graph low level questions nor to increase or decrease their rate of asking low level questions. An increase in high level questions did not appear to result in a corresponding decrease in low level questions, however.

Because there were no significant differences between the two experimental groups (E1 and E2) for lessons 8, 9 and 10, Boeck concluded that the treatment was equally effective regardless of the number of lessons taught before it was introduced. The treatment did not work equally well for all subjects. One subject showed no change in rate of asking high level questions or in the rate of pupil talk. This individual did not listen to his audiotapes or categorize his own questioning behavior as requested. Nine of the experimental subjects showed a substantial increase in the rate of high level questioning

after treatment and smaller or less stable increases were observed for the other four experimental subjects.

Blosser (1970) investigated the effectiveness of an instructional procedure designed to develop skill in questioning. As a part of this study, a question category system focused specifically on types of questions likely to be asked in science lessons was developed: Question Classification System for Science (QCSS). In addition to attempting to increase the number of open questions asked, students were also asked to increase the length of time they paused after asking a question and before accepting a student response, and to decrease the amount of teacher talk in their lessons.

Forty-two preservice secondary school science teachers were involved in this study. The study extended over several quarters but the treatment was the same for each group, each quarter. Students were randomly assigned to one of four groups. Three of the four groups were asked to microteach a 15-minute lesson to peers which was videotaped. Subjects in group R₁ were asked to read a handbook on questioning designed by the investigator and to serve as teacher for three microteaching lessons of the teach-analyze-reteach variety. Lesson 1 focused on use of open and closed questions and methods for decreasing teacher talk. Lesson 2 focused on open questions and on pausing. Lesson 3 focused on combining all three strategies. Three weeks were used for the treatment period, with the teach-reteach sessions for a particular lesson occurring within the same week.

Subjects in group R₂ served as students for the microlessons, participating in a teach or reteach session each week and evaluating the teacher's performance on the specific strategies to be demonstrated in the lesson. Subjects in group R₃ participated only in the pre- and posttest measures (microteaching lessons). Subjects in R₄ were post-tested only. Students were also observed during their student teaching quarter during their first week of teaching, at midquarter, and during the final week of student teaching, to see if any of the skills developed during instruction transferred to student teaching.

These lessons were audiotaped for data analysis. Transcripts were made from the tapes (audio, video) for use in data analysis by three judges who listened to randomly selected 15-minute segments of the student teaching lessons (to correspond in length with the 15-minute microteaching lessons). Rater reliability was determined by Guilford's method of interclass correlation and was .73 for category I of the QCSS, .70 for category II, and .70 for category III for the judges and .92 for category I, .91 for category II, and .90 for category III for the investigator. The judges did not know to which of the four groups the particular student whose questions they were categorizing belonged.

Blosser found that pausing behavior could be developed through practice and experience ($p = .01$) and that teacher talk could be decreased ($p < .01$). There was no significant change in the rate of asking open questions. Members of group R₁ were able to continue to use pausing techniques and to decrease teacher talk during student teaching but not at a level of significance (.10).

Concentrating on the same three questioning strategies (asking open questions, pausing, and decreasing teacher talk), Blosser conducted a second investigation (1972) with college juniors. She was interested in determining the combined effects of duration and timing of instruction in questioning techniques. Twenty-seven preservice secondary school science teachers were involved in the two-quarter study.

Subjects were randomly assigned to one of three treatment groups. Group R_1 received instruction in questioning distributed over a 14-week period across the two quarters. Groups R_2 and R_3 received the same instruction during a seven-week period. R_3 received the instruction during quarter one and R_2 , during quarter two. During quarter one the preservice teachers worked as tutors and teachers' assistants in junior high schools. During quarter two they worked, in teams of two, in elementary schools and taught science lessons to small groups of children.

Data were gathered by audiotaping lessons during quarter two. Blosser found no significant differences (.05 level) between treatment groups for the variables of open questions, pause time mean, and teacher talk. Neither the massed practice of five group sessions within a seven-week period nor the more distributed practice of five group sessions spread over a 14-week period differed in effectiveness.

When Blosser looked for a significant trend over time in developing questioning skill, she found no significant trend in asking open questions. When the use of pause time was considered, based on data from R_2 , it appeared that providing individuals with some instruction in a particular skill concurrent with expecting them to demonstrate some competency in that skill was the least effective situation. The trend over time for teacher talk was a linear one, indicating a decrease in teacher talk. However, it was not possible to determine if this decrease was due to instruction or to change in pupil population (peers to elementary school pupils).

Blosser reported that the results of this second study supported providing questioning instruction in a distributed fashion and then holding an audiotaping session with follow-up activities for maximum change. She speculated that if this were done in cycles, with a different major emphasis in each cycle, it might be possible to get the maximum gain for the skill being emphasized as well as perhaps providing an effect needed to increase retention of all the other questioning behaviors.

Adhikary (1973) evaluated instruction in questioning in terms of the effectiveness of bringing about a change in the question-asking behavior of prospective volunteer teachers. Thirty Peace Corps trainees were the subjects of this study. All were recent graduates of American universities with a major or minor in science and/or mathematics and were preparing to teach in Nepal. Adhikary used the Aschner-Gallagher system for classifying questions: cognitive-memory, convergent, divergent, evaluative, and managerial.

The Peace Corps trainees were pretested by having them teach a 15-minute lesson, to peers, in Napali. The lesson was audiotaped and the category and frequency of questions recorded. Instruction for the 15 individuals in the treatment group consisted of asking them to write 15 classroom questions representative of science or mathematics in Napali, pretesting them on classifying questions, and then providing them with programmed instruction on classifying teacher questions into categories. After completing the programmed instruction, the students were posttested on classifying questions. They then categorized and criticized the 15 questions they had written. This was followed by a discussion on the importance of asking different types of questions. Trainees were encouraged to write questions as part of lesson planning for their next peer teaching.

Posttesting consisted of a 30-minute lesson (audiotaped) of peer teaching in Napali. A 15-minute segment of this lesson was analyzed. Adhikary reported no significant differences in frequency of questions asked in each category between the control and experimental groups on the pretest. When pre- and posttest results on classifying questions were considered for the experimental group, there was a significant difference (.005) in scores in favor of the posttest. Apparently the training program did help the subjects to classify questions.

When performance of experimental and control groups on the posttest lesson was analyzed, Adhikary found no significant differences in the frequency of cognitive-memory questions, evaluative questions, and managerial questions. There was a significant difference (.01) in the frequency of convergent questions in favor of the experimental group and a significant difference (.05) in the frequency of divergent questions in favor of the experimental group. There was also a significant difference (.05) in the frequency of total questions in favor of the experimental group. Adhikary reported that both groups asked more questions on the posttest than they did on the pretest.

Another study completed in 1973 at Indiana University, as Adhikary's had been, was that of W. S. Smith. Smith reported that previous work by Ladd and Andersen had provided evidence of a strong relationship between teachers' question asking and student achievement. Therefore, preservice teachers should be helped with their use of high inquiry questions. Smith investigated the effects of alternative types of instruction on learning a question-asking skill by preservice secondary school science teachers.

Forty-six students were placed into four groups: a placebo group, which Smith identified as treatment 1; treatment group 2, which was given instruction on classifying questions; treatment group 3, which was given instruction on how to sequence high inquiry questions; and treatment group 4, which was given instruction on classifying questions on a test.

Smith differentiated between low inquiry questions and high inquiry questions. According to Smith, low inquiry questions include defining,

describing, designating, substituting, and reporting questions. High inquiry questions include comparing, explaining, conditional inferring, and evaluating questions (1973, p. 6).

All subjects were pretested and provided a 10-minute audiotape of peer microteaching in which they attempted to demonstrate good questioning techniques. Posttest I was another tape of a 10-minute microteaching lesson with peers, including a demonstration of skills learned during instruction. Posttest II was a 10-minute segment of a lesson taught during student teaching. Smith performed the analysis of the tapes, determining his reliability with Scott's reliability coefficient (.90 on pretest, .89 on posttest I, and .92 on posttest II).

Smith reported finding no significant differences among the groups in the use of high inquiry questions during the pretest, posttest I, and posttest II. Nor were there any significant differences among the groups in the use of high inquiry questions during posttest I or during posttest II, adjusted for the use of high inquiry questions during the pretest. The use of high inquiry questions decreased for all groups between posttest I and posttest II (1973, p. 38).

Coyne (1974) examined microteaching as a vehicle for acquiring a teaching skill: questioning. Coyne's problem was to study the effects of microteaching with instructor-prepared lessons compared with trainee-prepared lessons. Coyne's subjects were 24 preservice physics and mathematics teachers at the University of Cape Coast, Ghana.

The students were placed in two experimental groups and a control group by random assignment. One experimental group received microteaching training in questioning skills using supervisor-outlined lessons (SL group) while the other experimental group microtaught using lessons which they had prepared to develop skill in questioning (TL group). The control group did not microteach. Both experimental groups taught five 5-minute lessons.

Coyne reported that the questioning skills emphasized were the use of probing questions, question fluency, the use of higher order questions, and divergent questions. Appendix C in his dissertation contains three rating scales: fluency and probing questions, higher order and divergent questions scale, and questioning skills scale (1974, pp. 54-56). The rater is asked to respond to eight questions (nine on the higher order and divergent questions scale), using a 1-5 rating for each question with a rating of 5 being high.

Three judges used these scales to analyze the preservice teachers' questioning behaviors. Rater reliabilities are reported (0.80, 0.82 and 0.93) for the pooled-judges' ratings on the three rating scales although method for determining reliability is not specified. In addition to the tapes of the microteaching lessons, two lessons were taperecorded for all students during their student teaching.

Coyne reported no significant differences between the SL and TL groups for microteaching lessons emphasizing fluency and probing

questions. When higher-order and divergent questions were considered, the trainees who received lesson outlines demonstrated a superior performance, with a significant difference being identified. The ratings of questioning skills were significantly higher for the two treatment groups than for the control group for the lessons recorded during student teaching. There were no significant differences between the SL and TL groups on the student teaching tapes.

According to Coyne, the students in the SL group had significantly more positive attitudes toward microteaching than did those in the TL group. Both microteaching groups had significantly more positive attitudes toward the methods course than did the control group students who had experienced the usual methods course.

Lamb, in an article in Science Education (1977), reported on his work on questioning skill development involving four populations and three designs for the various investigations. The first study, which involved student teachers, will be discussed in this section of the review and the remaining studies will be described in the section of the review dealing with research involving inservice teachers.

Lamb developed, as a part of his dissertation research, a self-instructional module on questioning skill development. This module was developed after reviewing questioning research literature to determine those components of questioning skill training systems most likely to result in real and lasting changes in teachers' in-class questioning behavior (1977, p. 30). Lamb developed a five-category system for classifying questions (four of the five categories have two or three subcategories). He also developed a written programmed instructional sequence of 120 frames aimed at helping teachers learn to use the question category system. After completing the written program, trainees used audiotaped programmed instruction of teaching episodes chosen to demonstrate models of the desired overall questioning behavior as well as examples of specific question types. Lamb reported that time to complete the module averaged six and one-half hours, with a range of four and one-half hours to twelve and one-fourth hours.

In the studies designed to assess the effectiveness of the module, participants provided an audiotape of their science teaching prior to treatment and after treatment. Lamb and an assistant performed the tape analysis. Reliabilities for raters were established with Scott's Pi Coefficient (0.89 for Lamb and 0.86 for the assistant). Lamb hoped to find an increase in cognitive variety of questioning after treatment.

The experiment which involved preservice teachers also involved their cooperating teachers. Subjects were divided into three groups. In Group 1 both student teachers and cooperating teachers worked through the module. In Group 2 only the student teachers used the module. In Group 3 neither student teachers nor cooperating teachers used the module.

Ten-minute segments of three different lessons (taught to the same class) were randomly chosen for analysis. Lamb reported that the two groups of student teachers who worked through the module asked a wider variety of questions following treatment than prior to it. This held true when data were analyzed for the three groups with pretreatment questioning variety and supervising teacher feedback were held constant via ANCOVA.

Another study completed at The University of Texas, as Lamb's had been, was that of Crawley and Krockover (1979). As Lamb had done, they designed and tested the effectiveness (both immediate and delayed) of an instructional module whose goal was to improve the cognitive level of preservice secondary science teachers' questions. Their module, "Levels of Questioning," included a statement of objectives, a pre-assessment (Form A of the Test of Inquiry Processes), an activity sequence, and a posttest.

The pretest for the module (not the training sequence) involved the classification of ten questions according to the levels of Bloom's taxonomy. If nine of the ten questions were correctly classified, students worked on other tasks.

Within the module activities, students identified and then classified a set of ten test questions taken from two science textbooks (one traditional, one modern). They noted specific and general differences in the nature of the content of the questions. Follow-up discussions focused on the questioning levels identified in the first activity. Students were supplied with a chart of action verbs when asked to classify questions. They also made use of this chart when they prepared a set of ten test questions from their science area and grade level (for student teaching). Distribution of these questions across the levels of Bloom's taxonomy was specified: 20 percent from levels 1 and 2, 70 percent from levels 3 and 4, and 10 percent from levels 5 and 6. The instructors discussed and evaluated the question set with each preservice teacher ($N = 14$).

The posttest was designed to have the subjects demonstrate their ability to transfer ideas acquired from the module to classroom activities. Subjects also completed Form B of the Test of Inquiry Processes. The instructors gave each of the subjects a copy of their progress report which described the results of pre- and post-assessments.

Each subject also was required to develop a module as a part of the requirements of the methods course. The module was expected to contain the same components as the "Levels of Questioning" module: objectives, pretest, activities, and posttest.

Scores on the Test of Inquiry Processes, Forms A and B, as well as the student-developed module pretest questions were analyzed by the instructors who independently noted the level of questions. (The module pretest was considered to be a delayed posttest.) Immediate effects of the training sequence were highly significant. The average questioning level significantly increased from pretest to posttest

($p < .001$). The average questioning level for the delayed posttest was not significantly different from that of the immediate pretest, however (1979, p. 246). Crawley and Krockover considered that the results of significance tests indicated that their instructional module was effective in immediately raising the average question level but the gain was short-lived.

In addition, the investigators expressed concern because students completing the module appeared to over-emphasize the use of high-level questions. On the immediate posttests 13 percent of the questions were low level and 87 percent high level. The authors considered equal numbers of low level and high level questions the most desirable state of affairs.

In Summary

The nine studies involving preservice secondary school science teachers in the improvement of their questioning behavior are less uniformly positive than were those involving elementary teachers. It does appear possible to help preservice teachers increase the number of divergent and evaluative questions they ask (Konetski); the use of precision teaching (Boeck) appears one method for improving questioning behavior provided the subjects involved participate as expected. Students can learn to ask open questions but not always at a level of significance as well as increasing their pause/wait-times and decreasing teacher talk (Blosser) and distributed rather than massed practice appears preferable (Blosser).

Students can learn to classify questions but this skill does not always result in a significant change in the types of questions they use in teaching (Adhikary). Training in the use of high-inquiry questions does not always carry over into student teaching (Smith) although skills gained during microteaching using instructor-developed lesson outlines appears to show some immediate benefit (Coyne).

Modules designed to improve questioning behavior have been developed (Lamb, Crawley and Krockover) and these appear to produce a change in the cognitive variety of questions trainees ask (Lamb) but this change may be short-lived (Crawley and Krockover).

Experimental Studies Involving Pre-Service College Science Teachers

Two investigations were placed in this category (Murphy, 1972; A. B. Smith, 1977). Both involved teaching associates in college science classes.

Murphy (1972) investigated the effectiveness of certain aspects of an inservice training program with inexperienced graduate teaching assistants assigned to college freshman chemistry classes. Seventeen graduate

teaching associates were involved. These persons were stratified by relative rank on chemistry department placement exams: nine were ranked as high ability and eight, low ability. They were randomly assigned to either the experimental or the control group.

A pre-treatment observation of each subject was made and audio-taped. Subjects in the experimental group participated in an inservice program that consisted of classroom activities, individual conferences with the investigator, and microteaching lessons involving three to four freshman students. These microteaching sessions were videotaped and viewed with the investigator to discuss questioning skills. Inservice activities amounted to approximately one hour per week for seven weeks.

Two observations were made after the inservice sessions; one the following week and the second, nine weeks later. Murphy analyzed the questions asked, using the Question Category System for Science, reporting a reliability of .80. Murphy reported that so few divergent thinking questions and evaluative thinking questions were asked that these categories were not even included in his data analysis.

Murphy found that, in both post-treatment and late measures, TAs from the treatment group lectured less and received more student response. People in the treatment group also had more silence in their classes, on the late measure. There were no significant changes in the questioning behaviors of the subjects. The inservice program did not appear to have produced TAs who asked a greater variety of questions.

A. B. Smith (1977) wanted to promote the use of higher order questions asked by teaching associates working with college physics classes. Eighteen teaching associates were videotaped, for 15 minutes each, as they worked in an open lab. The tapes were rated by a person using the Florida Taxonomy of Cognitive Behavior System (FTCB).

The inservice program for the physics TAs consisted of five classroom hours of lecture and discussion over a five-week period and 12 observer ratings of each TA by another TA over the remaining four weeks of the term. One seminar included a discussion of various types of questions lab instructors might ask. (The Gallagher-Aschner system of cognitive-memory, convergent thinking, divergent thinking, and evaluative thinking questions was used.) TAs had the opportunity to practice levels of questioning in small groups. During the final week of the quarter, each TA has his behavior in the lab again videotaped for 15 minutes. These tapes were rated by the same individual who had analyzed the first tapes.

There were problems with the taping system and only 10 of the 18 tapes were good quality enough to permit analysis. However, Smith concluded that the objective of increasing higher level questioning by the TAs was achieved. He concluded that he did not achieve his second objective: increasing the proportion of higher level questions

by the TAs. He does report that prior to training the proportion of higher level questions was 54 percent. (Lamb and Crawley and Krock-over might speculate that the desired mix of high level and low level questions already existed, prior to instruction.)

In Summary

Murphy did not achieve an increase in variety of questions with chemistry TAs but Smith met a similar objective with physics TAs. Both used an inservice program with their subjects. They used different instruments to analyze questioning behavior but category II of the QCSS (Murphy) consists of the four categories of the Gallagher-Aschner system which Smith used.

Table 2

Experimental Studies of Questioning Behavior Involving Inservice Science Teachers

ELEMENTARY

Bedwell, 1974
Beisenherz, 1971
Johnson, 1969

SECONDARY

Aagaard, 1972
Hoosein, 1972
Lamb, 1977

Experimental Studies Involving Inservice Elementary School Science Teachers

Johnson (1969) wanted to develop a model program for improving the questioning behavior of inservice elementary school teachers in the area of science instruction. He worked with five teachers who had been enrolled in a six-week summer science institute, funded by NDEA. (There had been 88 participants in the workshop; five volunteered to participate in Johnson's study.) Johnson had established six criteria for participation in his research: there were videotape data on the teacher's behavior from the summer institute, the participants had worked within the same summer institute workshop group, they were still teaching at the primary level in 1968-69, they had taught a minimum of five years, they were fully certified with the minimum of a B.A. in elementary education, and they were willing to cooperate within the structured research program.

Johnson developed a question hierarchy based on the type of pupil response expected. From lowest to highest level, questions were categorized as routine-management, memory, observational, information-processing, and evaluative. Teachers were taught to plot the questions they asked on a Questioning Dissector Grid. The abscissa on the grid instrument was numbered serially from 0 to 90 and represented the natural sequence of questions asked by the teacher during a single lesson. The question categories were on the ordinate axis of the grid.

Johnson had prepared graphed data profiles for each of the five volunteers in his study from their summer workshop activities recorded on videotape. Each of the five teachers taped three lessons (each 20 minutes in length) during the study. Each teacher and the researcher categorized the questions asked in these lessons and plotted all questions sequentially on the Questioning Dissector Grid, according to their individual perceptions. Counts of numbers of questions asked and falling into each question category were tabulated and the tables given to each teacher to analyze her performance.

Two to three days after each recorded lesson, the researcher held individual post-lesson conferences with the participating teachers in their respective schools. During these conferences they compared category placement of questions, evaluated any gain over the previous lesson in level of productive thought responses solicited from pupils by each question, and focused on strategies of asking questions. The researcher attempted to assume a guidance role in these conferences.

Johnson hoped to achieve an increase in process-oriented questions (observational, information-processing, evaluative). He reported that four of the five teachers showed an increase in process-oriented questions asked (from the baseline data) over the three lessons. The participants asked a greater percentage of questions in the process-oriented categories as the study progressed. The percentage of memory-recall questions decreased for four of the five teachers.

Johnson concluded that lesson format, more than the teacher, controlled the questioning style (1969, p. 82). The format most likely to produce process-oriented questions was developmental in nature and characterized by student-centered investigations. Johnson thought that he could identify some questioning patterns emerging from the data, such as focus, lift, extend.

Beisenherz (1971) was involved with measuring the utilization and effectiveness of a locally televised elementary school science series, grades 1-4, in selected classrooms in the Seattle metropolitan area. Four areas of interest were identified. Two of these related to questioning: the analysis of television teachers' and classroom teachers' questions in terms of the types of questions asked, the context in which they were asked, and their location in the sequence of instruction selected for investigation and the analysis of the extent to which the quality and sequence of the questioning behavior of the studio teacher transferred to the questioning behavior of the classroom teacher.

The study involved four studio teachers and 54 teachers in 27 elementary schools. Pre and post measures were obtained in this study. All teachers taught the ESS Mystery Powders and provided an audiotape as the posttest measure. Two question category systems were used: a modification of the Gallagher-Aschner system (the category of "observation question" was inserted between cognitive-memory questions and convergent thinking questions) and an Instructional Strategy Category System (ISCS). The ISCS had four categories of questions: exploration, invention, discovery, and review.

Beisenherz reported that, in general, a higher proportion of convergent questions--in comparison with other question types--was asked by the television teachers and by classroom teachers using the televised science program. There were significant differences in memory, observation, and convergent questions between treatment and post-treatment periods. There was a higher proportion of convergent and observation questions and a lower proportion of memory questions for the post-treatment period.

Bedwell (1974) who felt that teachers needed to use questions to stimulate students' thinking designed a study to determine if teachers trained in question-asking skills would cause greater cognitive achievement and more positive attitudes among pupils they instructed using a high level questioning strategy than among pupils they instructed using a low level questioning strategy. Nine elementary school teachers and their fourth- and fifth-grade pupils were involved in this study (four fourth-grade classes, three fifth-grade classes, and two combined fourth- and fifth-grade classes).

All teachers provided a pre-treatment tape of 15-20 minutes of class discussion. The teachers participated in a five-session workshop. Questions were divided into high level (above Bloom's comprehension level) and low level (below Bloom's application level). Teachers used a self-instructional module on questioning consisting of a booklet and an audiotape. The module involved classifying written questions, using Bloom's taxonomy; planning and writing questions according to cognitive level; wait-time and sequencing; use of observations from the audiotape to classify "verbal questions" (1974, p. 23); and planning and teaching several practice lessons which included at least one question from each cognitive level so that 40 percent of the questions were above the comprehension level.

Bedwell analyzed data from three post-treatment observations. He concluded that teachers could be trained to write and classify questions according to cognitive level and that they could raise the cognitive level of class discussions via question-asking skills. However, three of the nine teachers were unable to ask 40 percent of the questions above the comprehension level, and two other teachers taped only two lessons rather than the requested three.

There were no significant differences on student achievement for high and low level question strategies nor were there any significant

differences on student attitude related to questioning strategy. Bedwell reported a "... slight trend in favor of the low level questioning strategy..." (1974, p. 62) and speculated that teachers may have been attempting to lay a knowledge base or have been handicapped in their questioning by the Piagetian stage of cognitive development of their students.

In Summary

Inservice elementary school teachers can change their questioning behavior--sometimes. Johnson reported that the participants in his study responded favorably to the post-lesson conferences which he held with each of them. Beisenherz reported great individual variation among teachers, with some teachers asking a disproportionate number of questions. He also noted that availability of science equipment and materials for use by classroom teachers was a factor in how successfully they implemented and extended the televised science program. Bedwell was not able to achieve results at a level of significance. Perhaps asking teachers to assume two different modes of questioning with randomly selected groups of students within their classroom handicapped the transfer of skill gained in writing and classifying questions to oral questioning behaviors in science lessons.

Experimental Studies Involving Inservice Secondary School Science Teachers

Aagaard (1972) identified what he considered to be two basic needs in chemistry education: to apply appropriate results of research in pedagogical techniques of questioning to accomplish maximum achievement through the use of high level questions and to encourage thought and participation in the learning process. In keeping with his beliefs, Aagaard investigated the possible effect of the use of selected question types on student achievement in eleventh-grade chemistry classes.

Aagaard established a set of criteria for participation and chose 14 chemistry teachers for his study. Ten were involved in the two experimental groups and four in a control group. The teachers taught a two-week unit on radiation chemistry, using different methods. The control group taught in an expository manner. One experimental group was given a procedural section for the unit containing 350 low level questions which might be used. The other experimental group received a procedural section containing 210 high level questions. The teachers provided tapes of six hours of classroom interaction so Aagaard could verify that they were following the instructional pattern he had designated.

Students were pretested before beginning the unit and took an achievement test at the end of the unit. Aagaard found no significant differences among student groups in pretest means. There was a significant difference ($p < .01$) among student groups in posttest means.

The difference was between experimental groups, not between experimental and control groups. The experimental group taught by high level questions had significantly higher achievement scores than did the group taught with low level (cognitive memory) questions or than did the control group.

Hoosein (1972) was also interested in determining if changes in teacher behavior can produce changes in student behavior. He was particularly interested in student inquiry as measured with the VIB instrument. This category system for interaction analysis consists of three categories for student behavior, six categories for teacher behavior, and two additional categories (silence, confusion). It is similar to the Flanders system, differing in that the student behavior categories come before those for teacher behavior and in that silence and confusion are two separate categories.

The 38 teachers involved in Hoosein's study participated in an inservice program with 24 receiving instruction for three hours each Wednesday after school in the Instructional Behavior and Skills Development Program for 36 weeks. The teachers supplied post-treatment audiotapes and 30 minutes of verbal interaction were coded by 10 persons for analysis by a computer program.

Although the inservice program emphasized questioning as a technique to increase student inquiry (1972, p. 90) and skill in classifying teacher questions, Hoosein reported that teacher questions were not modified. Within the experimental group, there were observable differences in both student and teacher talk categories, however. All student talk categories showed an increase in percentage of occurrence and teacher talk decreased.

Lamb (1977) used his self-instructional module on questioning with inservice teachers as well as in the study, described earlier in this review, with preservice teachers. In his second study, Lamb worked with 11 inservice teachers in an urban area. Five of the teachers worked through the self-instructional module and six served as a control group. Pre- and post-instructional audiotapes of lessons were analyzed (randomly selected 15-minute segments). Lamb found no significant differences between groups.

In the third study, eight science teachers from small rural schools completed the module as a requirement for a graduate course in science education. The fourth study was similar to number three but it involved six rural school secondary science teachers. These six teachers were volunteers and were not enrolled in any graduate class. The graduate students required to work through the module asked a wider variety of questions following treatment than prior to it ($p < .03$), based on t-test comparisons of pre- and post-treatment variety scores. There was no significant difference between pre- and post-treatment variety scores for experiment four.

Lamb reported that, in the second experiment, two of the five teachers in the treatment group returned their questioning modules with

the seals intact, indicating that they had not used the module and were, in effect, "untreated." He reported, "... In this study, only those groups coerced into completing their module demonstrated a significant increase in the cognitive variety of questioning, enforcing the conclusion that the module should not be used with volunteers..." (1977, p. 38). (Usually researchers assume a positive Hawthorne effect with volunteers in a study but Lamb apparently had reason to feel differently.)

In Summary

The three studies in this group provide different points of view relative to questioning. Apparently when teachers are given questions at a certain level of complexity and ask these questions, their students achieve at a higher level than students of teachers who ask low level questions (Aagaard). A lengthy inservice program focused on verbal interaction and designed to promote student inquiry through teacher questioning did not produce a significant difference nor were teachers' questions modified (Hoosein). The inservice materials were copyrighted and not included in the dissertation. The VIB category system contains only one category for teacher questions. Therefore it is impossible to know if the investigator's description of the aim of the program is an accurate one or in what ways the program focused on teacher questioning. (Perhaps being involved in inservice every Wednesday from 4-7 P.M. produced a negative set in the teacher participants.)

Lamb found that his instructional module did help teachers ask a wider cognitive variety of questions, provided they really worked through the module! He suggested the module was necessary but not sufficient--coercion/persuasion was also important.

Experimental Studies Focused on Wait-Time

The four studies in this cluster deal with the variable that Rowe calls "wait-time," the time a teacher waits, or pauses, after asking a question and before accepting or demanding a student response (wait-time I) or after a student has responded before the teacher talks again (wait-time II) (Rowe, 1974). The four investigators are Fowler, 1974; Garigliano, 1972; Lake, 1973; and Moriber, 1971.

Moriber (1971) investigated how much wait-time college science instructors provided their students. He worked with three college faculty members who taught physical science for nonscience students. Moriber chose the participants after looking at student evaluation forms of the laboratory instruction and supervisor observation reports so that his group included one individual rated excellent; one, very good; and two, fair. Moriber defined "wait-time" in his study as that time between a teacher question and a student response or between a teacher question and the rephrasing of this question (which apparently did not receive a student response).

Moriber's subjects taped a lesson, then discussed it with the investigator. The instructors were advised to slow down, to allow more wait-time. Observations were made weekly for five weeks. Verbal interaction was analyzed with Flanders interaction analysis system.

Moriber found that the teacher rated as "best" allowed very little wait-time and had a general teaching pattern that was that of lecture. Teacher 3, rated as poorest, allowed the most wait-time and asked the most questions. However, a large number of these questions were of the yes/no variety. Moriber also found that when teachers reworded questions that received no response, the second question was not exactly the same as the original one so students wondered which question they were expected to answer. Moriber reported that the pattern exhibited by the teachers during the first lesson observed was still the dominant pattern after the fifth week of observation, despite four lengthy individual conferences.

Garigliano (1972) worked with 33 teachers but what he was interested in studying was the effect of wait-time on selected discourse variables (of pupils) in microtaught SCIS lessons. He looked at the mean number of words in student responses, the mean number of inflected student responses, the mean number of content-oriented student solicitations, the mean number of content-relevant student-student interactions, and the mean number of "I don't know" responses.

The 33 teacher participants taught in four different schools but all were working with a trial center for SCIS. The teachers were randomly assigned to one experimental and two control groups. (Twenty-one of the teachers were teaching SCIS for the first time and four were not trained to use SCIS.) The experimental group and control group I each made three microtapes of SCIS lessons involving four pupils for a lesson. The experimental group got feedback on the use of the wait-time variable and tapescripts of their microteaching. Control group II made one microtape. The two control groups got feedback only on the use of SCIS lessons and materials.

For the wait-time training, the teacher and researcher listened to the tape of the microteaching and followed the tapescript for 20 minutes. The researcher provided rough counts on wait-time (a) after teacher questions and (b) after student responses. The teachers were asked to mentally count 1001-1005 to increase their wait-time. Garigliano reported that, because of his schedule, two to six-week intervals occurred between microlessons.

Garigliano looked at different types of wait-time, changes in wait-time, and changes in student behaviors within and between student groups. He found no significant differences among the three groups, in baseline data, on the independent variables of wait-time. When Garigliano analyzed data from the third tapes, he found that the experimental group and the control groups differed significantly ($p < .05$) in wait-time. However, the experimental group never reached a wait-time of three to five seconds, never getting beyond slightly less than two seconds (which still exceeded the overall mean wait-time of the control groups by almost one second).

Garigliano reported that the greatest gains in wait-time by the experimental group were achieved in what Rowe calls wait-time II, after a student response. Garigliano speculated that classroom interaction may really be a contest for talk space with children playing the role of respondents. Teachers appeared unable to make large gains in wait-time after asking a question because children were more in control of this particular talk space than were the teachers. However, the wait-time after a student response seems to be under the control of the teacher, with the children waiting for additional teacher questions, directions or reactions (1972, p. 71).

Garigliano found negative and significant correlations between wait-time after a teacher question and the number of "I don't know" responses and the number of content-oriented student solicitations. However, the magnitude of differences between groups was too small to permit the investigation of the relationship of wait-time to these variables.

No statistically supported differences were found between the experimental and the control groups on any of the dependent variables. Differences in mean wait-times were probably too small to have had any appreciable effect on the outcome variables, Garigliano concluded.

Lake (1973) investigated the effect of the extension of wait-time on student inquiry behavior. He was interested in the effect of extended wait-time on cognitive complexity of student responses. Lake served as the teacher for his study, working with 72 fifth-grade students (18 groups of four each). Nine of these 18 groups were taught the SCIS material "Making Paper Airplanes," using a wait-time of one second or less and nine groups, a wait-time of three seconds or more. The lessons extended over a two-week period.

Lake considered that student inquiry was taking place if the student talk displayed speculation, conversational sequences, alternative explanations, and/or arguments over interpretation of data. Lessons were audiotaped and typescripts produced. Two judges were involved. The Bellack classification of the three modes in cognitive processes involved in classroom verbal interaction was chosen for data analysis. Only two of the three modes were used in Lake's study: the empirical mode, involving fact stating or explaining, and the evaluative mode, involving opining or justifying.

The discovery phase of the SCIS lessons was chosen for analysis. Inter-coder reliability was established in terms of percent of agreement achieved by coders on each item of typescript, as Bellack had specified.

Lake reported that the incidence of conversational sequences was significantly greater ($p < .05$) under long wait-time, as was the incidence of alternative explanations. Significant differences ($p < .05$) were found between the incidence of ascending modal congruence under long and short wait-time conditions. Lake concluded that extending wait-time did increase student inquiry behavior, and that under long wait-time conditions, student responses became cognitively more complex.

Lake suggested that wait-time be "re-defined" as "the length of the silent period which occurs immediately before a teacher utterance" (1973, p. 3). This would emphasize the need to wait before interjecting with teacher talk. Lake's remarks agree with those of Garigliano in that both consider the wait-time after a teacher's question to be under the control of the students rather than of the teacher.

Fowler (1974) also focused on wait-time. He was interested in documenting some of the behavioral and cognitive aspects of wait-time and in delineating whatever interrelationships might exist between the various forms of wait-time as well as in assessing the relationships that might exist between wait-time and teacher attitudes toward pupil control.

Fowler worked with preservice elementary school teachers enrolled in a science methods course. The methods course involved a three-week field experience in elementary school, four weeks of on-campus instruction, and six more weeks of field experience. Twenty of the 51 preservice teachers were selected for wait-time training. These 20 preservice teachers were randomly assigned to one of two groups just before they taught a lesson. One group was asked to use a wait-time of one-half second; the other, three seconds. Students were given a form of assistance in judging their wait-time by the method of transferring a cube from one hand to the other. Those requested to wait three seconds were told to transfer four cubes, one at a time. Those with the one-half second or no wait-time were told to transfer only one cube.

Students were given practice in using the two types of wait-time and also participated in a discussion of the problems of waiting in silence. For purposes of data analysis the preservice teachers each taught a group of four sixth- and seventh-grade students. Thirty-minute lessons were recorded and 10 minutes of each lesson were analyzed.

Attitudes toward Pupil Control were obtained by the use of the Pupil Control Ideology. Fowler found no significant differences relative to PCI scores and wait-time.

Fowler looked at the incidence of student-student interaction, student questions, students' suggestions for testing inferences, the coherence of group verbal interaction, and the proportion of verbal involvement for each student in the group. Fowler found that students whose teachers were asked to wait three seconds made 18 percent fewer inferences, about the same number of suggestions for testing inferences, asked about the same number of questions, initiated 28 percent more statements, verbally interacted with other students 41 percent more frequently than students in the no wait-time group. The three-second wait-time group also had 45 percent less interruptions and 33 percent fewer statements to teacher directives. He found no support for wait-time increasing the incidence of student questions, forming inferences or developing suggestions for testing inferences.

Fowler reported that the duration of wait-time in the wait-time group was about twice that in the no-wait group (1974, p. 53). The teacher reaction wait-time (TRWT), which took place after a student utterance and before a teacher utterance, had a mean value of less than one second. Even with the small value, Fowler found the incidence of student-student interactions increased and student inferences decreased.

In Summary

It would be difficult to generalize from these four studies of wait-time. Two involved inservice teachers (Moriber, Garigliano) but one group was teaching elementary students and the other, college students. Lake served as the teacher for his study and followed a series of carefully scripted lessons. Fowler worked with preservice teachers.

Both Garigliano and Lake do differentiate between student-controlled wait-time (Rowe's wait-time I) and teacher-controlled wait-time (Rowe's wait-time II). Fowler performed a factor analysis of his data and identified two factors: teacher-controlled silence and student-controlled silence, which appear to support the statements of Garigliano and Lake.

ADDITIONAL CONSIDERATIONS

So far in this review both descriptive studies of teachers' questioning behaviors in science classrooms and experimental studies designed to change teachers' questioning behaviors have been considered. The studies have been further grouped as dealing with preservice teachers or inservice teachers and further separated by educational level: elementary, secondary and college. Studies dealing with wait-time rather than question classification have also been included. Where possible, similarities have been identified or major points emphasized in summary sections.

In this section of the review the discussion will focus on such factors as the question category system involved, the length of instructional treatment, materials used in the treatment (if these are specified), and a brief review of the outcomes of the studies in an attempt to identify any factors which should be considered by future investigators of questioning skills, either for inclusion or exclusion in future research.

Question Category Systems Used

Just as the ten categories of the Flanders interaction analysis instrument have influenced other investigators of verbal interaction, the Gallagher-Aschner system of classifying questions appears to have influenced a number of investigators. This is evident in Table 3 (following page). Six studies used the Gallagher-Aschner system and a seventh included the Gallagher-Aschner categories within a three-level category system.

The four categories within the Gallagher-Aschner system: cognitive-memory, convergent thinking, divergent thinking, and evaluative thinking are based on the Guilford model of the structure of the intellect. This system is not so complex as to present difficulties to subjects who are expected to memorize and apply it in formulating questions. Its use enables investigators to lump categories, if this suits their purposes, and talk about low level questions (usually containing only those of the cognitive-memory variety) and high level questions.

Five investigators used Bloom's taxonomic categories of the cognitive domain for classifying questions. One of these five researchers (Boeck) used a three-category system: routine, rhetorical, and pedagogical. However, the pedagogical category was further divided into low level thought (memory-translation) and high level thought (application, analysis, synthesis, evaluation), basically Bloom's categories. Another of these researchers (Bozardt) reported the use of the system developed by Sanders and published in Classroom Questions: What Kinds? (1966). The system developed by Sanders is basically the levels of Bloom's taxonomy modified to make them useful to social studies teachers interested in improving their questioning skills.

Table 3

Studies Clustered by Question Category System

Gallagher-Aschner	Bloom	Question Category System for Science
Adhikary	Bedwell	Blosser***
Beisenherz	Boeck	Murphy
Cunningham	Bozardt*	Riley
Konetski	Bruce**	
Larson	Crawley & Krockover	
Masla		

*Used Sander's question hierarchy which is an adaptation of Bloom's taxonomy.

**Although questions were grouped as lower and higher level, the levels were those of Bloom's taxonomy with the addition of "demonstration of skill" as a higher level and without the "evaluation" category.

***Level II within this category system consists of the four categories of the Gallagher-Aschner system.

The Question Category System for Science (QCSS) was used by three researchers in four studies (two by the same investigator). This is a three-level system. Level I consists of closed questions (those with a limited number of acceptable responses), open questions (those with a greater number of acceptable responses), managerial questions, and rhetorical questions. Within Level II, closed questions are further classified as either cognitive-memory or convergent thinking and open questions, as either divergent thinking or evaluative thinking (the Gallagher-Aschner categories). Level III consists of further classification of each of the four categories of Level II. Persons learning to use this system are expected to develop proficiency with the use of Level I before proceeding to Level II and then to Level III. Based on reading the reports of research, the investigators stressed Levels I and II in their instruction.

Other systems were used, such as Parson's Self-Guided Analysis System (Anshutz) or Harris and McIntyre's Teacher Question Inventory (Porterfield). Parson's system classified questions as rhetorical, informational, leading and probing. The Harris and McIntyre system has two components. The "cognition" component resembles a modification of the Bloom taxonomy: recognition, recall, demonstration of skill, translation, interpretation, analysis and synthesis. The other component, "affectivity," is divided into opinion and attitudes or values.

In some reports, probably because of page limitations on journal articles, it was not possible to identify categories or subcategories. Koran wrote about observation and classification questions and was presumably operating within a two-category system. Rice divided questions into lower cognitive and higher cognitive, based on key words in question phrasing, but this system is not fully described in the journal article.

It would seem logical to assume that the fewer the categories that teachers need to learn to memorize, the higher the possibility that they would use the system. However, it is also logical to assume that other factors may be of more importance than are the length and complexity of the category system.

Length of Instructional Treatment

The length of the treatment for the experimental group was not always obvious from reading the research report or journal article. Frequently, number of sessions might be specified but not their length. In those studies in which treatment time could be identified or inferred with some degree of confidence, times appeared to range from 55 minutes (Santiesteban) to 36 weeks (Hoosein) for three hours per week. In several studies the subjects in the experimental group(s) went through a self-instructional module (Bozardt, Crawley and Krockover, Konetski, Lamb, Rice) but only Lamb reported the minimum, maximum, and average times needed to complete his module.

In some studies part of the treatment involved conferences with the researcher. The average length of these conferences was seldom specified.

Length of instructional treatment may, or may not, be a factor that is of major importance in designing a questioning study. However, if length of instructional treatment influences attitude of the participants, the researcher needs to keep in mind the other demands on his/her subjects' time. If the instructional treatment is built into a methods course or some inservice activity, chances are greater that it will be carried out in the manner the investigator wishes. Even then Boeck reported that not all of her subjects analyzed their lessons upon completion of each lesson (as desired) so that this feedback could be used when planning and teaching the next lesson. Lamb reported, in one study, that two of five participants did not even break the seal on the self-instructional questioning module he had designed and were, therefore, "untreated."

While all researchers might not agree with Lamb's remarks that subjects should be "coerced" into completing the treatment, as specified, it is essential that the subjects consider the activity important enough to complete it in the manner the investigator desires.

Materials Used in Treatment

Again these varied widely. Perhaps a reason for this wide variation is the fact that most of the research studies reviewed were doctoral dissertations and, as such, the emphasis is usually more on originality than on replication.

In some studies the problem was to test the effectiveness of a particular training technique. Frequently more than one outcome was desired from this technique. As a result the techniques and outcomes were varied.

In some instances the emphasis was placed on the use of models, either audio or video or both (Koran, Santiesteban, Rice). In others, self-instructional or programmed materials were involved (Bozardt, Konetski, Lamb, Bedwell, Crawley & Krockover).

The materials used and directions for their use were not always available, as an appendix, in the dissertation. In some instances copyrighted materials were involved and these were excluded from the microfilming.

In some studies the emphasis on question asking was a part of a workshop program and was apparently carried out in a relatively informal manner with more stress placed on learning to use science curriculum improvement materials than on classifying questions.

In two studies the materials involved scripted lessons into which the questions had been incorporated (Aagaard, Lake). In those studies

involving the investigation of the effects of wait-time, the emphasis was on recognizing appropriate points in the lesson to remain silent and learning to wait. Rice emphasized wait-time, along with other variables, in his study and had his subjects read Rowe's "Science, Silence and Sanctions" (1969). Fowler supplied his subjects with cubes to transfer. Garigliano suggested that his subjects mentally count 1001 . . . 1005 to increase their wait-times.

Outcomes of Studies

In some reviews of research there is a section which may be described as a "win-loss" section in which the reviewer designates those studies where significant differences were identified in favor of the experimental group (win column) and those in which no significant differences were found (loss column). However, such a tally does not really provide more information than that found in the descriptions of the studies and their findings.

Instead, the remarks in this section will concern the bases for the various investigations and some comments about techniques or procedures which appeared to have some degree of success in changing questioning behaviors of science teachers.

Bases for the Research

It was not always possible to determine, from reading the introductory portion of chapter one or from the grouping in the literature review chapter, the rationale underlying the research. However, more often than not, the investigator did provide some remarks which appeared to indicate why he or she decided to research a particular problem. In general, the researchers appear to hold the idea that teachers' questions influence students' thinking processes. (Probably if we did not believe this we would be hard put to explain why we don't concentrate on lecture and forget about inquiry teaching for science.)

Again, researchers appear to believe that inquiry teaching is not widely used or understood and that teachers need help in modifying their teaching behavior to make better use of inquiry. One way to make better use of inquiry is to ask, in Lamb's phraseology, a wider cognitive variety of questions. Researchers also appear to believe that asking "higher level" questions will help students develop the ability to think. In some respects this sounds a little like faculty psychology, now held in disrepute, but it more probably relates to the idea of transfer of training than to the exercise of a specific faculty.

If researchers believe that asking a wider cognitive variety of questions promotes student thinking and achievement, it would seem logical that they should test for this in their research. The majority of those persons conducting experimental studies did not do this.

Some did, however (e.g. Anshutz, Aagaard, Hoosein, Bedwell). Those who did not investigate the effects of changes in teacher questioning behavior on student behavior were probably operating in a first-things-first mode. They were primarily interested in learning if the sequence, self-instructional module, or activities could bring about a change in teacher questioning behavior. Once they determined this, the particular investigation or dissertation problem was completed.

Results of studies in which student achievement was considered or other student factors were involved are mixed. Aagaard found a significant difference in favor of the student group with which the teacher had used high level questions. Anshutz found no significant differences relative to student achievement or attitude. Bedwell found no significant difference in student achievement related to questioning strategy. In Bedwell's report he notes a "... slight trend in favor of low level questions..." (1974, p. 62) which, he speculated, may have been related to needing to provide a knowledge base or to Piagetian level of cognitive development of the students involved.

Those individuals focusing their research on wait-time provided varied reasons for their focus but, in general, these reasons related to a desire for more student participation or for the use of student inquiry. Again, not all researchers analyzed student behavior but the majority in this small sample (three of the four) of researchers did. The use of extended wait-time did produce significant changes in student behaviors (see study descriptions in an earlier section of this review).

Rowe has reported that, as teachers become comfortable with using wait-time, the number and kinds of questions they ask change. Total number of teacher questions decreases (students are talking longer) and the number of questions calling for reflection or clarification of meaning increases (1973, p. 261).

It would appear that those individuals concerned with increasing wait-time and its effect on student behavior were more likely to study this relationship than were those researchers who were concerned about the use of higher level questions by teachers in the hope that such usage would influence student thinking and achievement. Perhaps this occurred because helping teachers learn to pause or wait involves less complexity than helping them learn to differentiate question complexity, use variety in their question asking, and then study the use of this variety (if it was achieved) on the students of these teachers.

Even learning to increase the length of wait-time is not so easy as it may sound. Garigliano reported that the teachers in his experimental group never achieved the criterion length (3-5 seconds) for a silent period. He does report, however, that one teacher in the control group overheard a conversation in the teachers' lounge about his study and attempted to do what he (Garigliano) wished to have done: increase wait-time. This individual, in attempting to please the researcher or to work toward what he considered a desirable goal, had a longer average

wait-time than the teachers in the experimental group on the third lesson taped after treatment. So, we are brought back again to the issue of motivation or cooperation raised by Lamb as well as to the wide variability in individual data mentioned by Boeck, Johnson and others.

Are some individuals more motivated to change their behavior than others are? If so, could researchers identify such persons and concentrate their efforts with these subjects? Personality factors and results of the Pupil Control Ideology do not provide data for identifying such groups for questioning studies in the literature reviewed here.

Apparently Successful Techniques or Procedures

Although a number of the investigations resulted in findings at a level of statistical significance, there appear to be some techniques that are more likely to be successful than others. Those studies in which models were used (Koran, Santiesteban, Rice) contain findings that suggest subjects are more likely to master a skill if they can see or hear it modelled for them. Johnson reported that the non-directive, supportive conferences he held with his subjects as they discussed the feedback they obtained from graphing their questioning behavior were positively received. Blosser used individual conferences with subjects in her second study and reported that she could identify some changes, if only temporary, in questioning behavior after these conferences.

Boeck considered what she termed the "operant methodology" of micro-teaching to be a useful strategy in changing teacher behavior (1970, p. 62). In Boeck's study the subjects were expected to categorize and graph their questions upon completion of a lesson. Not all subjects did this, but those who complied with directions were able to obtain a type of concrete feedback concerning their behavior in the form of the graphed data. Johnson's Questioning Dissector Grid is another instance of this type of information. Perhaps such a visual representation of oral behavior is more helpful to teachers (preservice or inservice) than is a general discussion of questioning behavior exhibited in a particular lesson. The focus of the change, or lack of change, is probably sharper when the behavior is presented in visual form.

On the other hand, those investigations in which self-instructional modules were used or self-analysis was emphasized did not produce statistically significant results for the most part. When the control of the instructional treatment is given over to the subjects to be treated, there is no guarantee that the instruction will proceed as desired or even that it will take place (as shown in one of Lamb's studies). Reviews of research on auto-tutorial programs provide evidence that students in such programs have a tendency to procrastinate. This was true even in Boeck's study which did not involve self-instruction but which did involve self-analysis.

It would appear that designing an instructional treatment in which subjects hear (and, possibly, see) positive models of the behavior to be learned, in which they receive some graphic representation of their behavior, and in which they participate in supervisory conferences may result in desired changes in behavior.

IN CONCLUSION

The descriptive studies of science teachers' questioning behavior provide evidence that teachers ask a large number of questions and that the majority of these questions are at the cognitive-memory, or recall, level in terms of thinking operations required by students in order to respond. In that respect, science teachers are probably no different from teachers in other content areas in terms of the type of questioning behaviors they typically exhibit.

Numerous investigators have conducted studies designed to change the questioning behavior of teachers or their use of wait-time, or both. Some appear to have been more successful than others, in terms of significant findings reported.

In those investigations in which follow-up data have been gathered, the behavioral changes gained as a result of instructional treatment have, for the most part, been short-lived. Behaviors exhibited in micro-teaching sessions, usually involving peers, are not usually exhibited at a level of statistical significance during student teaching if pre-service teachers are involved. Studies of inservice teachers usually involve teachers working with public school pupils for the treatment, if this is, indeed, a part of the instructional treatment as well as during the lessons from which follow-up data are analyzed. The transition from a less-than-real-life setting of microteaching to the classroom is not present. In studies involving inservice teachers, skills do, for the most part, appear to transfer although not always at a level of significance.

The majority of the studies reviewed were doctoral dissertations. As such they are subject to the usual set of criticisms relative to design, appropriateness of statistical techniques, and generalizability of the findings. In some instances the investigators acknowledge some methodological flaws but others are unaware of such flaws or unwilling to acknowledge their awareness in print.

There appears to be little effort to build upon previous research in designing instructional sequences or using materials from other studies. Perhaps this may again be traced to the emphasis upon originality rather than on replication in doctoral studies. Perhaps the literature is not adequately reviewed before the investigation is begun. Or perhaps the materials from previous studies are not readily available for use by other investigators.

It does appear possible to change the questioning behaviors of teachers, either preservice or inservice, as well as to teach them to wait longer for student discourse. Perhaps the type of research now needed is that of using some of the previously developed materials and techniques with a greater emphasis on the effects of changed teacher behaviors on student behaviors.

Possible Research

Sufficient data are at hand to negate the statement that "little has been done" to change teachers' questioning behaviors. However, work needs to/could be done in several areas. The sequencing of questions and the effect of this on student responses has been little investigated in science. Are there certain patterns or sequences of questions that are more productive of student talk and thinking than others? Some work in this area was done by Hilda Taba and her colleagues with elementary school children in social studies lessons (1964).

The majority of the studies reviewed here concentrated on teachers' questions and neglected the manner in which teachers followed up student responses to their questions. Some investigation of this area has been done in those studies involving wait-time but not in any great detail.

Three of the four investigations concerning wait-time involved elementary school science. More research needs to be done about the effects of wait time on secondary school science teachers and their students. Will learning to wait (particularly before a teacher utterance, as Lake advocates) have the same effects on classroom verbal interaction and pupil participation for secondary school science classes as have been found in elementary school science lessons?

Some of the descriptive studies contained information about the influence of curriculum materials on teachers' questions as did the experimental study by Riley. Is it possible that the more traditional (whatever that term means) science programs and materials do not lend themselves to a variety of question types? Gall, writing in a survey of research on questions over a 50-year period (1970), suggests that researchers need to identify valued educational objectives in a specific setting and then construct questions to enable students to reach each objective (1970, p. 711). Perhaps asking teachers to change their questioning behavior without providing them materials to teach that are compatible with the desired question types is wasted effort.

Another point in Gall's survey report deserves emphasis here. Gall wrote, "It is important that teachers' questions should not be viewed as an end in themselves. They are a means to an end--producing desired changes in student behavior. . . ." (1970, p. 718). Materials have been developed for use by science teachers, as well as by teachers of other content areas. Preservice programs as well as inservice programs need

to be provided with models, specific suggestions, and with specific feedback including some in visual form if their behavior is to change. Although it is hoped that teachers will become self-analytical, they need supervisory feedback while questioning skills are being developed and behavior change is stabilized.

The assistance of Dr. Diana Stano in identifying sources and in the preliminary scanning of materials is appreciated.

REFERENCES

- Aagaard, Stanley A. "Oral Questioning by the Teacher: Influence on Student Achievement in Eleventh Grade Chemistry." Unpublished doctoral dissertation, New York University, 1972.
- Adhikary, Poona Kanta. "Instructional Effect on Question Asking Behavior of Prospective Peace Corps Science and Mathematics Volunteer Teachers." Unpublished doctoral dissertation, Indiana University, 1972.
- Anshutz, Ramona J. "An Investigation of Wait-Time and Questioning Techniques as an Instructional Variable for Science Methods Students Microteaching Elementary School Children." Ann Arbor, Michigan: University Microfilms, 1973. (University of Kansas)
- Balzer, A. L.; T. P. Evans; and P. E. Blosser. A Review of Research on Teacher Behavior. Columbus, Ohio: ERIC Information Analysis Center for Science, Mathematics and Environmental Education, December, 1973.
- Bedwell, Lance E. "The Effects of Training in Question-Asking Skills on the Achievement and Attitudes of Elementary Pupils." Ann Arbor, Michigan: University Microfilms, 1974. (Indiana University)
- Beisenherz, Paul C. "An Experimental Study of a Televised Science Series, Grades One-Four, Comparing the Quality and Sequence of Television and Classroom Questions with a Proposed Strategy of Science Instruction." Ann Arbor, Michigan: University Microfilms, 1971. (University of Washington)
- Beisenherz, Paul C. and Jerry L. Tucker. "Limitations of Research in Science Teacher Questioning Behavior." Paper presented at the annual meeting of the National Association for Research in Science Teaching, Detroit, Michigan, March, 1973.
- Blosser, Patricia E. "A Study of the Development of the Skill of Effective Questioning by Prospective Secondary School Science Teachers." Ann Arbor, Michigan: University Microfilms, 1970. (The Ohio State University)
- Blosser, Patricia E. "The Development of the Skill of Questioning in Prospective Secondary School Science Teachers: An Extension." Final Report, Project No. 2-E-010, Grant No. OEG-5-72-0043 (509). Washington, D.C.: U.S. Office of Health, Education and Welfare, National Center for Educational Research and Development, December, 1972.
- Boeck, Marjorie A. "Experimental Analysis of Questioning Behavior of Pre-Service Secondary School Science Teachers." Ann Arbor, Michigan: University Microfilms, 1970. (University of Minnesota)

- Bozardt, Delphine A. "Development of Systematic Questioning Skills in an Elementary Science Methods Course." Ann Arbor, Michigan: University Microfilms, 1974. (University of Georgia)
- Bruce, Larry R. "A Determination of the Relationships Among SCIS Teachers' Personality Traits, Attitude Toward Teacher-Pupil Relationship, Understanding of Science Process Skills and Question Types." Ann Arbor, Michigan: University Microfilms, 1969. (Michigan State University)
- Coyne, George W. "The Effectiveness of Supervisor Versus Trainee Prepared Lessons on the Acquisition of Selected Microteaching Skills." Ann Arbor, Michigan: University Microfilms, 1974. (Purdue University)
- Cunningham, Roger T. "A Descriptive Study Determining the Effects of a Method of Instruction Designed to Improve the Question-Phrasing Practices of Prospective Elementary Teachers." Unpublished doctoral dissertation, Indiana University, 1968.
- Crawley, Frank E. and Gerald H. Krockover, "Immediate and Delayed Effects of Training Preservice Secondary Science Teachers to Ask Questions of Varying Cognitive Levels." Journal of Research in Science Teaching 16(3): 243-248, 1979.
- Doyle, Patrick J. "An Investigation into Factors Affecting the Learning and Utilization of a Questioning Technique by Elementary Methods Students." Ann Arbor, Michigan: University Microfilms, 1972. (Indiana University)
- Fowler, Thaddeus William. "An Investigation of the Teacher Behavior of Wait-Time." Ann Arbor, Michigan: University Microfilms, 1974. (University of Houston)
- Gall, Meredith. "The Use of Questions in Teaching." American Educational Research Journal 40(5): 707-721, December, 1970.
- Garigliano, Leonard J. "The Relation of Wait Time to Student Behaviors in Science Curriculum Improvement Study Lessons." Ann Arbor, Michigan: University Microfilms, 1972. (Columbia University)
- Grunau, Harold H. "An Experimental Study of the Effects of a Preservice Elementary Science Curriculum Innovation on Selected Measures of Teacher and Pupil Performance." Ann Arbor, Michigan: University Microfilms, 1972. (Michigan State University)
- Hoeteker, James and William P. Ahlbrand, Jr. "The Persistence of the Recitation: A Review of Observational Studies of Teacher Questioning Behavior." Bethesda, Maryland: ERIC Document Reproduction Service, 1968. ED 036 511.
- Hoosein, Abdool N. "The Effect of Teacher Behavior Modification Training upon the Level of Student Inquiry." Ann Arbor, Michigan: University Microfilms, 1972. (University of Missouri-Columbia)

- Hunter, Elizabeth. "The Effects of Training in the Use of New Science Programs upon the Classroom Verbal Behavior of First Grade Teachers as They Teach Science." Classroom Interaction Newsletter IV, pp. 5-11, May, 1969.
- Johnson, Robert W. "A Model for Improving In-Service Teacher Questioning Behavior in Elementary School Science Instruction." Ann Arbor, Michigan: University Microfilms, 1969. (Wayne State University)
- Kleinman, Gladys S. "General Science Teachers' Questions, Pupil and Teacher Behaviors, and Pupils' Understanding of Science." Ann Arbor, Michigan: University Microfilms, 1964. (University of Virginia)
- Kondo, Allan K. "The Questioning Behavior of Teachers in the Science Curriculum Improvement Study Teaching Material Objects." Paper presented at the National Association for Research in Science Teaching annual meeting, Pasadena, California, February, 1969.
- Konetski, Louis C. "Instructional Effect on Questions Asked by Pre-service Science Teachers." Ann Arbor, Michigan: University Microfilms, 1969. (Indiana University)
- Koran, John J., Jr. "A Comparison of the Effects of Observational Learning and Self-Rating on the Acquisition and Retention of a Questioning Behavior by Elementary Science Teacher Trainees." Science Education 54(4): 385-389, October-December, 1970.
- Koran, John J., Jr. "The Relative Effects of Classroom Instruction and Subsequent Observational Learning on the Acquisition of Questioning Behavior by Pre-Service Elementary Science Teachers." Journal of Research in Science Teaching 6(3): 217-223, 1969.
- Lamb, William G. "Evaluation of a Self-Instructional Module for Training Science Teachers to Ask a Wide Cognitive Variety of Questions." Science Education 61(1): 29-39, 1977.
- Lake, John H. "The Influence of Wait-Time on the Verbal Dimension of Student Inquiry Behavior." Ann Arbor, Michigan: University Microfilms, 1973. (Rutgers University)
- Larson, Charles G. "An Analysis of Classroom Questioning Strategies in Elementary Science." Ann Arbor, Michigan: University Microfilms, 1973. (Memphis State University)
- McGlathery, Glenn. "Analyzing the Questioning Behaviors of Science Teachers," in What Research Says to the Science Teacher, Volume I, M. B. Rowe (ed.). Washington, D.C.: National Science Teachers Association, 1978.
- Masla, John A. "The Effects of Instruction in Interaction Analysis on the Verbal Inquiry Patterns of Elementary Science Methods Students." Ann Arbor, Michigan: University Microfilms, 1968. (Indiana University)

- Moon, Thomas C. "A Study of Verbal Behavior Patterns in Primary Grade Classrooms During Science Activities." Ann Arbor, Michigan: University Microfilms, 1969. (Michigan State University)
- Moriber, George. "Wait-Time in College Science Classes." Science Education 53(3): 321-328, 1971.
- Moyer, John R. "An Exploratory Study of Questioning in the Instructional Processes in Selected Elementary Schools." Ann Arbor, Michigan: University Microfilms, 1965.
- Murphy, Michael D. "The Development and Assessment of an Experimental Teacher Training Program for Beginning Graduate Assistants in Chemistry." Ann Arbor, Michigan: University Microfilms, 1972. (The Ohio State University)
- Porterfield, Denzil R. "Influence of Preparation in Science Curriculum Improvement Study on Questioning Behavior of Selected Second and Fourth Grade Reading Teachers." Ann Arbor, Michigan: University Microfilms, 1969. (University of Oklahoma)
- Rice, Dale R. "The Effect of Question-Asking Instruction on Preservice Elementary Science Teachers." Journal of Research in Science Teaching 14(9): 353-359, 1977.
- Riley, Joseph P., II. "Effects of Studying a Question Classification System on the Cognitive Level of Preservice Teachers' Questions." Science Education 62(3): 333-338, 1978.
- Rowe, Mary Budd. "Relation of Wait-Time and Rewards to the Development of Language, Logic, and Fate Control, Part II: Rewards." Journal of Research in Science Teaching 11(4): 291-308, 1974a.
- Rowe, Mary Budd. "Science, Silence, and Sanctions." Science and Children VI, pp. 11-13, March, 1969.
- Rowe, Mary Budd. Teaching Science as Continuous Inquiry. New York: McGraw-Hill Book Co., 1973.
- Rowe, Mary Budd. "Wait-Time and Rewards as Instructional Variables, Their Influence on Language, Logic, and Fate Control, Part I: Fate Control." Journal of Research in Science Teaching 11(2): 81-94, 1974b.
- Santiesteban, A. Joseph. "The Relative Effects of Video and Audio Models on the Acquisition of a Teaching Skill by Preservice Elementary Teachers and Concomitant Student Learning." Ann Arbor, Michigan: University Microfilms, 1974. (University of Florida)
- Smith, Albert B. et al. "Improving Instruction by Measuring Teacher Discussion Skills." American Journal of Physics 45(1): 83-87, January, 1977.

Smith, Walter S. "The Effect of Alternative Types of Instruction on the Learning of a Question-Asking Skill by Preservice Secondary Science Teachers." Ann Arbor, Michigan: University Microfilms, 1973. (Indiana University)

Stewart, Martin D. "Cognitive and Affective Process Development and Their Relations to the Uses of Lecture and Transitions Among Lecture, Questions, and Student Initiated Comments." Ann Arbor, Michigan: University Microfilms, 1974. (New York University)

Taba, Hilda; Sam Levine; and F. F. Elzey. Thinking in Elementary School Children. Cooperative Research Project No. 1574. San Francisco: San Francisco State College, 1964.

Wilson, John H. "Differences Between Inquiry-Discovery and the Traditional Approaches to Teaching Science in the Elementary Schools." Ann Arbor, Michigan: University Microfilms, 1969. (University of Oklahoma)